Supplement of Earth Syst. Dynam., 17, 495–532, 2026
https://doi.org/10.5194/esd-17-495-2026-supplement

Earth System Dynamics
Open Access

Supplement of

Developing Guidelines for working with Multi-Model Ensembles in CMIP

Anja Katzenberger et al.

Correspondence to: Anja Katzenberger (anja.katzenberger@gmx.de)

S1 Statistics of the field over past decades

Figures 4 and 5 were built using data from the Web of Science database. The queries for each category are:

Total ML:

TS=("machine learning" OR "artificial intelligence" OR "neural networks" OR "random forest" OR "decision trees" OR "deep learning" OR "supervised learning" OR "unsupervised learning") AND TS=("CMIP" OR "CMIP3" OR "CMIP5" OR "CMIP6" OR "Coupled Model Intercomparison Project" OR "climate model" OR "climate models" OR "general circulation model" OR "general circulation models" OR "Earth system model" OR "Earth system models")

ML-MME:

TS=("machine learning" OR "artificial intelligence" OR "neural networks" OR "random forest" OR "decision trees" OR "deep learning" OR "supervised learning" OR "unsupervised learning") AND TS=("CMIP" OR "CMIP3" OR "CMIP5" OR "CMIP6" OR "Coupled Model Intercomparison Project" OR "climate model" OR "climate models" OR "general circulation model" OR "general circulation models" OR "Earth system model" OR "Earth system models") AND TS=("multi-model ensemble" OR" multi-model ensembles")

ML-Downscaling:

TS=("machine learning" OR "artificial intelligence" OR "neural networks" OR "random forest" OR "decision trees" OR "deep learning" OR "supervised learning" OR "unsupervised learning") AND TS=("CMIP" OR "CMIP3" OR "CMIP5" OR "CMIP6" OR "Coupled Model Intercomparison Project" OR "climate model" OR "climate models" OR "general circulation model" OR "general circulation models" OR "Earth system model" OR "Earth system models") AND TS=("downscaling" OR "bias correction")

ML-Downscaling MME:

TS=("machine learning" OR "artificial intelligence" OR "neural networks" OR "random forest" OR "decision trees" OR "deep learning" OR "supervised learning" OR "unsupervised learning") AND TS=("CMIP" OR "CMIP3" OR "CMIP5" OR "CMIP6" OR "Coupled Model Intercomparison Project" OR "climate model" OR "climate models" OR "general circulation model" OR "general circulation models" OR "Earth system model" OR "Earth system models") AND TS=("downscaling" OR "bias correction") AND TS=("multi-model ensemble" OR" multi-model ensembles")

ML Causality:

TS=("CMIP" OR "CMIP3" OR "CMIP5" OR "CMIP6" OR "Coupled Model Intercomparison Project" OR "climate model" OR "climate models" OR "general circulation model" OR "general circulation models" OR "Earth system model" OR "Earth system models") AND TS=("causal discovery" OR "causality" OR "causal inference" OR "causal")

ML Emulators:

TS=("machine learning" OR "artificial intelligence" OR "neural networks" OR "random forest" OR "decision trees" OR "deep learning" OR "supervised learning" OR "unsupervised learning") AND TS=("CMIP" OR "CMIP3" OR "CMIP5" OR "CMIP6" OR "Coupled Model Intercomparison Project" OR "climate model" OR "climate models" OR "general circulation model" OR "general circulation models" OR "Earth system model" OR "Earth system models") AND TS=("emulation" or "surrogate" or "emulator" or "emulators" or "surrogates")

ML XAI:

TS=("machine learning" OR "artificial intelligence" OR "neural networks" OR "random forest" OR "decision trees" OR "deep learning" OR "supervised learning" OR "unsupervised learning") AND TS=("CMIP" OR "CMIP3" OR "CMIP5" OR "CMIP6" OR "Coupled Model Intercomparison Project" OR "climate model" OR "climate models" OR "general circulation model" OR "general circulation models" OR "Earth system model" OR "Earth system models") AND TS=("XAI" OR "explainable AI" OR "Layer-wise Relevance Propagation" OR "LRP" OR "Feature importance analysis" OR "feature importance")

Model Independence:

TS=("climate" OR "Earth" OR "Earth System") AND TS=("CMIP" OR "Coupled Model Intercomparison Project" OR "climate model" OR "general circulation model") AND TS=("ensemble" OR "multi-model ensemble") AND TS=("dependence" OR "independence" OR "genealogy")

SMILEs:

TS=("Multi-model ensemble" OR "coupled model intercomparison" OR "cmip") AND TS= ("large ensemble" OR "grand ensemble" OR "smile")

S2 Systematic Model Biases

Some systematic biases are present in the vast majority of CMIP models at the global and regional scale, and some might even persist over multiple CMIP generations, which requires special attention. In this section we review some long-standing biases in CMIP models and strive to discuss their origins and consequences of these systematic model biases. Those are as follows: a) General evaluation, b) Sea surface temperature (SST) and ocean model biases, c) The Intertropical Convergence Zone (ITCZ) bias, d) biases in extratropical cyclones, e) Marine tropical/subtropical low cloud biases, f*)* biases in the cryosphere, *g)* biases in extremes. With this list, we do not intend to provide a complete list of all bias reported, but to give some relevant examples of model biases and its background. For further details on this topic, we also recommend Simpson et al. (2025).

General evaluation: Bock et al., 2020 employed the ESMValTool (see Section 2.6 and Eyring et al., 2020; Righi et al., 2020), to quantify the progress of climate models across different CMIP phases. Their analysis revealed significant advancements from CMIP3 to CMIP6 in simulating the vertical distributions of key variables, including temperature, water vapor, and zonal wind speed. The authors also demonstrated that high-resolution models in the historical CMIP6 simulations show a notable reduction of temperature and precipitation mean biases.

Sea surface temperature and ocean model biases: The ocean accumulates more than 90% of the excess energy from the global greenhouse effect (IPCC, AR6). The oceanic global circulation gyres transport excess heat from the tropics towards the poles. Furthermore, the oceanic surface fluxes of heat and moisture enter the atmosphere and thereby affect its dynamics. The ocean component also interacts with the cryosphere and influences processes therein (IPCC, AR6). These various oceanic processes have to be properly captured in ESMs. Long-standing SST biases result in biases when simulating other key phenomena such as tropical cyclones (e.g. Dutheil et al., 2020) and extratropical cyclones (e.g., Priestley et al., 2023a). Wills et al. (2022) investigated systematic biases in the large-scale patterns of recent SST and sea-level pressure change and showed that CMIP5 and CMIP6 ensembles are not able to reproduce the observed trends. Luo et al. (2023), moreover, discussed the origins of Southern Ocean warm SST bias in CMIP6 models. The Southern Ocean has namely been subjected to systematic warm SST bias in several generations of CMIP models (Sen Gupta et al., 2009; Wang et al., 2014). Westen and Dijkstra (2024) recently discussed persistent climate model biases in the Atlantic Ocean's freshwater transport. These various aforementioned biases are linked to the Atlantic Meridional Overturning Circulation (AMOC), which consists of the northward flow in the upper oceanic layers and returning southward flow in the deep ocean (Luo et al., 2023; Wang et al., 2024). The AMOC is considered to be one of the major tipping elements in the global climate system (Armstrong McKay et al., 2022; Van Westen et al., 2024), which may weaken or even collapse with future global warming, thus a more reliable representation of SST/ocean model would be desirable e.g. to better foresee the future AMOC behaviour.

The Intertropical Convergence Zone (ITCZ) bias: ITCZ is a band of a zonally-oriented surface convergence zone near the equator associated with deep convective clouds and heavy precipitation (Schneider et al., 2014; Waliser and Gautier, 1993). The common problem of fully-coupled global climate models from the early stage of their development is that they simulate two ITCZs over the central and eastern Pacific and the Atlantic in both hemispheres, instead of one ITCZ over the northern hemisphere as in observations, which is referred to as the double-ITCZ bias (Adam et al., 2018; Li and Xie, 2014; Oueslati and Bellon, 2015; Tian and Dong, 2020; Xiang et al., 2017). Tian and Dong (2020), as an illustration, recently examined the double-ITCZ bias in CMIP3, CMIP5, and CMIP6 based on annual mean precipitation. They found that all three generations of CMIP models exhibit similar systematic annual MME mean precipitation errors in the tropics when evaluated against the NOAA Global Precipitation Climatology Project (GPCP; Adler et al., 2003) and the NASA Tropical Rainfall Measurement Mission (TRMM; Huffman et al., 2007) observational datasets.

Biases in extratropical cyclones: Extratropical cyclones involving weather fronts and related overall storm tracks are an important component of the climate system since they transport heat poleward and are associated with a notable amount of precipitation and severe weather in the midlatitudes (Clark and Gray, 2020; Dacre, 2020; Schultz et al., 2019). The accurate representation of extratropical cyclones, including their thermodynamics, frontal structure, and track in CMIP models, however, remains challenging and has been subjected to biases (e.g. Chang et al., 2012; Priestley et al., 2023a, b). Priestley et al. (2023a) investigated drivers of biases in the CMIP6 extratropical storm tracks in the Northern Hemisphere (NH). Even though the previous work demonstrated that the representation of extratropical storm tracks in the NH has improved from CMIP5 to CMIP6, the persistent biases remain in CMIP6 (Priestley et al., 2023a). A follow-up study by Priestley et al. (2023b) investigated drivers of biases in the CMIP6 extratropical storm tracks in the Southern Hemisphere (SH). The Southern Hemisphere storm tracks have been commonly simulated too far equatorward in CMIP models during the historical period. This issue was somewhat reduced in CMIP6 compared to CMIP5, although it is still a problem.

Marine tropical/subtropical low cloud biases: Črnivec et al. (2023) analyzed 12 CMIP6 ESMs and demonstrated that they all underestimate the aerial extent of low clouds and simultaneously overestimate their radiative effect at the top of the atmosphere. This well-known issue, referred to as the “too few, too bright” tropical low-cloud bias, was already present in previous generations of climate models such as CMIP5 and CMIP3 (e.g., Nam et al., 2012, and references therein). Cesana et al. (2023), moreover, addressed how the representation of marine tropical Sc and Cu clouds and associated feedbacks in the abrupt 4xCO2 scenario changed between CMIP5 and CMIP6. They found that, collectively, CMIP6 models notably increased Sc cloud cover and slightly increased Cu cloud cover compared to their CMIP5 predecessors and are thus closer to observations. They further showed that CMIP6 models notably improved the representation of Sc feedback and slightly improved the representation of Cu feedback compared to CMIP5 models. Yet CMIP6 models still underestimate the magnitude of positive Sc and Cu feedbacks relative to observationally inferred estimates, which should drive further climate model development.

Biases in the cryosphere: The global cryosphere plays an important role in determining the planetary climate since bright ice and snow surfaces reflect a significant portion of the solar radiation back to space and cool the planet (IPCC, AR6). In a warming world, sea ice is shrinking and thinning, with both Arctic and Antarctic sea ice approaching historic lows (NASA Earth Observatory; IPCC AR6). The melting of sea ice with global surface warming implies that an increasing area of dark and absorptive ocean surface is exposed to warming sunlight, which forms one of the principal climate feedback mechanisms – namely, the sea ice albedo feedback (IPCC, AR6). It is thus pivotal to best capture the cryosphere extent, properties, and its response to global warming. To that end, Frankignoul et al. (2024) investigated Arctic September sea ice concentration biases in CMIP6 models and their relationships with other model variables. They demonstrated that CMIP6 models exhibit large biases in Arctic sea ice climatology, which seem to be related to biases in seasonal oceanic and atmospheric circulations. Notz and the Sea-Ice Model Intercomparison Project (SIMIP) Community (2020) furthermore showed that CMIP6 models still fail to simulate a plausible evolution of Arctic sea-ice area (SIA), even though CMIP6 models better capture the sensitivity of Arctic sea ice to forcing changes compared to CMIP5 and CMIP3 models. Roach et al. (2020) evaluated the Antarctic sea ice in CMIP6 and demonstrated that the mean Antarctic sea-ice area is close to satellite observations, but inter-model spread remains substantial, with summer Antarctic SIA being consistently biased low across the ensemble. Nevertheless, they found modest improvements in the simulation of sea-ice area and concentration compared to CMIP5.

S3 Further Examples of process-based Evaluation

Another example are low-level clouds over tropical and subtropical oceans that have been poorly simulated in multiple CMIP generations when evaluated against satellite observations in the present-day climate (e.g. Nam et al., 2012), which inhibits reliable future climate projections. Črnivec et al. (2023) and Cesana et al. (2023) introduced a qualitative approach to discriminate stratocumulus (Sc) from shallow cumulus (Cu) low-cloud regimes to evaluate their horizontal extent (cloud cover), radiative effect at the top of the atmosphere (TOA) and cloud-radiative feedbacks in CMIP5 and CMIP6 models. This approach is essential for guiding model improvements, because Sc and Cu formation and evolution are driven by a distinct interplay of coupled processes within the moist marine boundary layer (such as radiation, turbulence, convection); and Sc and Cu clouds also respond differently to global warming (Cesana and Del Genio, 2021).

The teleconnection between the Indian Summer Monsoon (ISM) and the El Niño–Southern Oscillation (ENSO) serves as a further example, which is well captured by MMEs of CMIP5 and CMIP6 models (Roy and Tedeschi, 2016; Roy et al., 2017). The teleconnection is strongest over central northeast India (Roy et al., 2017), where El Niño events are associated with a significant rainfall deficit, while La Niña events lead to a significant rainfall excess. In these studies, the MME is constructed using a simple mean ("one-model-one-vote") approach. Similar results are obtained when the MME is restricted to a subset of well-performing models, as identified for the ISM by Jourdain et al. (2013). Precipitation anomalies associated with different ENSO phases are reproduced well by most individual models and by the MME, in agreement with observations (see Roy et al., 2017 for details). Furthermore, the model ensemble analysis of ISM precipitation and Pacific SSTs reveals a clear linkage between the Walker circulation and the ISM over central northeast India, consistent with observations. This region represents the meeting point of the Hadley and Walker circulations during the ISM season, and the associated coupling and teleconnection processes appear to be well represented in most CMIP models and in the MME. This process-based understanding helps explain why the ENSO–ISM teleconnection is robustly captured in climate models.

Ahmed and Neelin (2021) utilised the observed relationship between tropical precipitation and buoyancy as the basis for a process-oriented analysis of CMIP6 models. They quantitatively assessed the thermodynamic sensitivity of convection across models using regime-oriented diagnostics. Their results showed that several models exhibit excessive moisture sensitivity, potentially arising from underactive convective schemes or tuning assumptions. Consequently, these models tend to produce mean precipitation states that are biased towards grid-scale saturation.

S4 Storylines for understanding Uncertainty

Considering this, uncertainty in climate projections can be communicated through climate storylines (Shepherd et al., 2018), which emphasizes exploring and understanding physically plausible events or pathways. The storyline approach differs from traditional methods of uncertainty evaluation in climate models in that it does not assume that model spread adequately represents uncertainty. This assumption may not hold for dynamically driven climate phenomena, where MME means may obscure regional details with individual climate models exhibiting atmospheric circulation patterns that can differ qualitatively from the multi-model mean (Bellomo et al., 2021; Zappa and Shepherd, 2017). Instead of quantifying the likelihood of events, storylines focus on the physical drivers and interactions that make an event possible (Shepherd et al., 2018), constructing a causal network and conditioning on specific assumptions. If we know thermodynamic changes are robust, the thermodynamic aspects of the observed changes are regarded as certain and the dynamic aspects as uncertain. By explicitly linking causal mechanisms to regional climate hazards, storylines are especially useful for regional climate impacts and understanding extreme events (Bevacqua et al., 2022; Shepherd, 2019; Zappa and Shepherd, 2017), improving the interpretability and usability of projections for decision-makers (Kunimitsu et al., 2023).

References

Adam, O., Schneider, T., and Brient, F.: Regional and seasonal variations of the double-ITCZ bias in CMIP5 models, Clim Dyn, 51, 101–117, https://doi.org/10.1007/s00382-017-3909-1, 2018.

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E.: The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present), J. Hydrometeor, 4, 1147–1167, https://doi.org/10.1175/1525-7541(2003)004%253C1147:TVGPCP%253E2.0.CO;2, 2003.

Ahmed, F. and Neelin, J. D.: A Process-Oriented Diagnostic to Assess Precipitation-Thermodynamic Relations and Application to CMIP6 Models, Geophysical Research Letters, 48, e2021GL094108, https://doi.org/10.1029/2021GL094108, 2021.

Armstrong McKay, D. I., Staal, A., Abrams, J. F., Winkelmann, R., Sakschewski, B., Loriani, S., Fetzer, I., Cornell, S. E., Rockström, J., and Lenton, T. M.: Exceeding 1.5°C global warming could trigger multiple climate tipping points, Science, 377, eabn7950, https://doi.org/10.1126/science.abn7950, 2022.

Bellomo, K., Angeloni, M., Corti, S., and von Hardenberg, J.: Future climate change shaped by inter-model differences in Atlantic meridional overturning circulation response, Nat Commun, 12, 3659, https://doi.org/10.1038/s41467-021-24015-w, 2021.

Bevacqua, E., Zappa, G., Lehner, F., and Zscheischler, J.: Precipitation trends determine future occurrences of compound hot–dry events, Nat. Clim. Chang., 12, 350–355, https://doi.org/10.1038/s41558-022-01309-5, 2022.

Bock, L., Lauer, A., Schlund, M., Barreiro, M., Bellouin, N., Jones, C., Meehl, G. A., Predoi, V., Roberts, M. J., and Eyring, V.: Quantifying Progress Across Different CMIP Phases With the ESMValTool, JGR Atmospheres, 125, e2019JD032321, https://doi.org/10.1029/2019JD032321, 2020.

Cesana, G. V. and Del Genio, A. D.: Observational constraint on cloud feedbacks suggests moderate climate sensitivity, Nat. Clim. Chang., 11, 213–218, https://doi.org/10.1038/s41558-020-00970-y, 2021.

Cesana, G. V., Ackerman, A. S., Črnivec, N., Pincus, R., and Chepfer, H.: An observation-based method to assess tropical stratocumulus and shallow cumulus clouds and feedbacks in CMIP6 and CMIP5 models, Environ. Res. Commun., 5, 045001, https://doi.org/10.1088/2515-7620/acc78a, 2023.

Chang, E. K. M., Guo, Y., and Xia, X.: CMIP5 multimodel ensemble projection of storm track change under global warming, J. Geophys. Res., 117, 2012JD018578, https://doi.org/10.1029/2012JD018578, 2012.

Clark, P. A. and Gray, S. L.: Sting jets in extratropical cyclones: a review, Quart J Royal Meteoro Soc, 146, 1065–1065, https://doi.org/10.1002/qj.3761, 2020.

Črnivec, N., Cesana, G., and Pincus, R.: Evaluating the Representation of Tropical Stratocumulus and Shallow Cumulus Clouds As Well As Their Radiative Effects in CMIP6 Models Using Satellite Observations, JGR Atmospheres, 128, e2022JD038437, https://doi.org/10.1029/2022JD038437, 2023.

Dacre, H. F.: A review of extratropical cyclones: observations and conceptual models over the past 100 years, Weather, 75, 4–7, https://doi.org/10.1002/wea.3653, 2020.

Dutheil, C., Lengaigne, M., Bador, M., Vialard, J., Lefèvre, J., Jourdain, N. C., Jullien, S., Peltier, A., Sultan, B., and Menkès, C.: Impact of projected sea surface temperature biases on tropical cyclones projections in the South Pacific, Sci Rep, 10, 4838, https://doi.org/10.1038/s41598-020-61570-6, 2020.

Eyring, V., Bock, L., Lauer, A., Righi, M., Schlund, M., Andela, B., Arnone, E., Bellprat, O., Brötz, B., Caron, L.-P., Carvalhais, N., Cionni, I., Cortesi, N., Crezee, B., Davin, E. L., Davini, P., Debeire, K., De Mora, L., Deser, C., Docquier, D., Earnshaw, P., Ehbrecht, C., Gier, B. K., Gonzalez-Reviriego, N., Goodman, P., Hagemann, S., Hardiman, S., Hassler, B., Hunter, A., Kadow, C., Kindermann, S., Koirala, S., Koldunov, N., Lejeune, Q., Lembo, V., Lovato, T., Lucarini, V., Massonnet, F., Müller, B., Pandde, A., Pérez-Zanón, N., Phillips, A., Predoi, V., Russell, J., Sellar, A., Serva, F., Stacke, T., Swaminathan, R., Torralba, V., Vegas-Regidor, J., Von Hardenberg, J., Weigel, K., and Zimmermann, K.: Earth System Model Evaluation Tool (ESMValTool) v2.0 – an extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP, Geosci. Model Dev., 13, 3383–3438, https://doi.org/10.5194/gmd-13-3383-2020, 2020.

Frankignoul, C., Raillard, L., Ferster, B., and Kwon, Y.-O.: Arctic September Sea Ice Concentration Biases in CMIP6 Models and Their Relationships with Other Model Variables, Journal of Climate, 37, 4257–4274, https://doi.org/10.1175/JCLI-D-23-0452.1, 2024.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., and Stocker, E. F.: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales, Journal of Hydrometeorology, 8, 38–55, https://doi.org/10.1175/JHM560.1, 2007.

Intergovernmental Panel on Climate Change (IPCC): Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, https://doi.org/10.1017/9781009157896, 2021.

Jourdain, N. C., Gupta, A. S., Taschetto, A. S., Ummenhofer, C. C., Moise, A. F., and Ashok, K.: The Indo-Australian monsoon and its relationship to ENSO and IOD in reanalysis data and the CMIP3/CMIP5 simulations, Clim Dyn, 41, 3073–3102, https://doi.org/10.1007/s00382-013-1676-1, 2013.

Kunimitsu, T., Baldissera Pacchetti, M., Ciullo, A., Sillmann, J., Shepherd, T. G., Taner, M. Ü., and van den Hurk, B.: Representing storylines with causal networks to support decision making: Framework and example, Climate Risk Management, 40, 100496, https://doi.org/10.1016/j.crm.2023.100496, 2023.

Li, G. and Xie, S.-P.: Tropical Biases in CMIP5 Multimodel Ensemble: The Excessive Equatorial Pacific Cold Tongue and Double ITCZ Problems*, Journal of Climate, 27, 1765–1780, https://doi.org/10.1175/JCLI-D-13-00337.1, 2014.

Luo, F., Ying, J., Liu, T., and Chen, D.: Origins of Southern Ocean warm sea surface temperature bias in CMIP6 models, npj Clim Atmos Sci, 6, 1–8, https://doi.org/10.1038/s41612-023-00456-6, 2023.

Nam, C., Bony, S., Dufresne, J. -L., and Chepfer, H.: The 'too few, too bright' tropical low-cloud problem in CMIP5 models, Geophysical Research Letters, 39, 2012GL053421, https://doi.org/10.1029/2012GL053421, 2012.

Oueslati, B. and Bellon, G.: The double ITCZ bias in CMIP5 models: interaction between SST, large-scale circulation and precipitation, Climate Dynamics, 44, 585–607, https://doi.org/10.1007/s00382-015-2468-6, 2015.

Priestley, M. D. K., Ackerley, D., Catto, J. L., and Hodges, K. I.: Drivers of Biases in the CMIP6 Extratropical Storm Tracks. Part I: Northern Hemisphere, Journal of Climate, 36, 1451–1467, https://doi.org/10.1175/JCLI-D-20-0976.1, 2023a.

Priestley, M. D. K., Ackerley, D., Catto, J. L., and Hodges, K. I.: Drivers of Biases in the CMIP6 Extratropical Storm Tracks. Part II: Southern Hemisphere, Journal of Climate, 36, 1469–1486, https://doi.org/10.1175/JCLI-D-20-0977.1, 2023b.

Righi, M., Andela, B., Eyring, V., Lauer, A., Predoi, V., Schlund, M., Vegas-Regidor, J., Bock, L., Brötz, B., De Mora, L., Diblen, F., Dreyer, L., Drost, N., Earnshaw, P., Hassler, B., Koldunov, N., Little, B., Loosveldt Tomas, S., and Zimmermann, K.: Earth System Model Evaluation Tool (ESMValTool) v2.0 – technical overview, Geosci. Model Dev., 13, 1179–1199, https://doi.org/10.5194/gmd-13-1179-2020, 2020.

Roach, L. A., Dörr, J., Holmes, C. R., Massonnet, F., Blockley, E. W., Notz, D., Rackow, T., Raphael, M. N., O'Farrell, S. P., Bailey, D. A., and Bitz, C. M.: Antarctic Sea Ice Area in CMIP6, Geophysical Research Letters, 47, e2019GL086729, https://doi.org/10.1029/2019GL086729, 2020.

Roy, I. and Tedeschi, R.: Influence of ENSO on Regional Indian Summer Monsoon Precipitation—Local Atmospheric Influences or Remote Influence from Pacific, Atmosphere, 7, 25, https://doi.org/10.3390/atmos7020025, 2016.

Roy, I., Tedeschi, R. G., and Collins, M.: ENSO teleconnections to the Indian summer monsoon in observations and models, Intl Journal of Climatology, 37, 1794–1813, https://doi.org/10.1002/joc.4811, 2017.

Schneider, T., Bischoff, T., and Haug, G. H.: Migrations and dynamics of the intertropical convergence zone, Nature, 513, 45–53, https://doi.org/10.1038/nature13636, 2014.

Schultz, D. M., Bosart, L. F., Colle, B. A., Davies, H. C., Dearden, C., Keyser, D., Martius, O., Roebber, P. J., Steenburgh, W. J., Volkert, H., and Winters, A. C.: Extratropical Cyclones: A Century of Research on Meteorology's Centerpiece, Meteorological Monographs, 59, 16.1-16.56, https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0015.1, 2019.

Sen Gupta, A., Santoso, A., Taschetto, A. S., Ummenhofer, C. C., Trevena, J., and England, M. H.: Projected Changes to the Southern Hemisphere Ocean and Sea Ice in the IPCC AR4 Climate Models, Journal of Climate, 22, 3047–3078, https://doi.org/10.1175/2008JCLI2827.1, 2009.

Shepherd, T. G.: Storyline approach to the construction of regional climate change information, Proc. R. Soc. A., 475, 20190013, https://doi.org/10.1098/rspa.2019.0013, 2019.

Shepherd, T. G., Boyd, E., Calel, R. A., Chapman, S. C., Dessai, S., Dima-West, I. M., Fowler, H. J., James, R., Maraun, D., Martius, O., Senior, C. A., Sobel, A. H., Stainforth, D. A., Tett, S. F. B., Trenberth, K. E., Van Den Hurk, B. J. J. M., Watkins, N. W., Wilby, R. L., and Zenghelis, D. A.: Storylines: an alternative approach to representing uncertainty in physical aspects of climate change, Climatic Change, 151, 555–571, https://doi.org/10.1007/s10584-018-2317-9, 2018.

Simpson, I. R., Shaw, T. A., Ceppi, P., Clement, A. C., Fischer, E., Grise, K. M., Pendergrass, A. G., Screen, J. A., Wills, R. C. J., Woollings, T., Blackport, R., Kang, J. M., and Po-Chedley, S.: Confronting Earth System Model trends with observations, Sci. Adv., 11, eadt8035, https://doi.org/10.1126/sciadv.adt8035, 2025.

Tian, B. and Dong, X.: The Double-ITCZ Bias in CMIP3, CMIP5, and CMIP6 Models Based on Annual Mean Precipitation, Geophysical Research Letters, 47, e2020GL087232, https://doi.org/10.1029/2020GL087232, 2020.

Van Westen, R. M., Kliphuis, M., and Dijkstra, H. A.: Physics-based early warning signal shows that AMOC is on tipping course, Sci. Adv., 10, eadk1189, https://doi.org/10.1126/sciadv.adk1189, 2024.

Waliser, D. E. and Gautier, C.: A Satellite-derived Climatology of the ITCZ, Journal of Climate, 6, 2162–2174, https://doi.org/10.1175/1520-0442(1993)006%253C2162:ASDCOT%253E2.0.CO;2, 1993.

Wang, C., Zhang, L., Lee, S.-K., Wu, L., and Mechoso, C. R.: A global perspective on CMIP5 climate model biases, Nature Clim Change, 4, 201–205, https://doi.org/10.1038/nclimate2118, 2014.

Wang, S., Sankaran, S., and Perdikaris, P.: Respecting causality for training physics-informed neural networks, Computer Methods in Applied Mechanics and Engineering, 421, 116813, https://doi.org/10.1016/j.cma.2024.116813, 2024.

van Westen, R. M. and Dijkstra, H. A.: Persistent climate model biases in the Atlantic Ocean's freshwater transport, Ocean Science, 20, 549–567, https://doi.org/10.5194/os-20-549-2024, 2024.

Wills, R. C. J., Dong, Y., Proistosecu, C., Armour, K. C., and Battisti, D. S.: Systematic Climate Model Biases in the Large-Scale Patterns of Recent Sea-Surface Temperature and Sea-Level Pressure Change, Geophysical Research Letters, 49, e2022GL100011, https://doi.org/10.1029/2022GL100011, 2022.

Xiang, B., Zhao, M., Held, I. M., and Golaz, J.: Predicting the severity of spurious "double ITCZ" problem in CMIP5 coupled models from AMIP simulations, Geophysical Research Letters, 44, 1520–1527, https://doi.org/10.1002/2016GL071992, 2017.

Zappa, G. and Shepherd, T. G.: Storylines of Atmospheric Circulation Change for European Regional Climate Impact Assessment, Journal of Climate, 30, 6561–6577, https://doi.org/10.1175/JCLI-D-16-0807.1, 2017.